

What is claimed is:

1. 1. A method for using an audio input from a telephony device to perform an action on an Internet protocol ("IP") network, the method comprising:
  3. providing a telephony interface module;
  4. receiving at the telephony interface module from the telephony device a first packet signal conforming to a telephony packet protocol and having an audio portion;
  6. receiving at the telephony interface module from a second module in communication with the telephony interface module (i) a second packet signal conforming to an IP, the second packet signal having an audio portion and (ii) a command;
  9. routing the first packet signal in accordance with the received command;
  10. converting, in the telephony interface module, the second packet signal to a third packet signal conforming to a telephony packet protocol and including an audio portion; and
  12. transmitting the third packet signal to the telephony device.
1. 2. The method of claim 1 further comprising:
  2. routing the first packet signal to a navigation module in communication with the telephony interface module;
  4. converting, in the navigation module the audio portion of the first packet signal to a text equivalent signal;
  6. converting, in the telephony interface module, the text equivalent signal to an IP network command signal; and
  8. using the IP network command signal to retrieve a document from the IP network.
1. 3. The method of claim 2 wherein the retrieved document is a voice XML document from the Internet.

1    4.     The method of claim 2 wherein the retrieved document is an HTML document from the  
2     Internet.

1    5.     The method of claim 4 wherein the second module is a text to speech module, the method  
2     further comprising:

3            receiving a displayable text portion of the HTML document;

4            converting the displayable text portion to an equivalent audio signal and converting the  
5     audio signal to an IP-based packet signal, thereby generating the second IP packet signal.

1    6.     The method of claim 1 wherein the step of receiving at the telephony interface module  
2     from the telephony device further comprises using a telephony gateway to convert an audio  
3     signal from a circuit switched signal to the first packet signal conforming to a telephony packet  
4     protocol and having an audio portion.

1    7.     The method of claim 1 wherein the step of transmitting the third packet signal to the  
2     telephony device further comprises using a telephony gateway to convert the third packet signal  
3     to a circuit switched signal thereby generating an audio signal receivable by the telephony device  
4     over the PSTN.

1    8.     The method of claim 1 wherein the telephony packet protocol conforms to one of a H.323  
2     and a SIP communications standard.

1    9.     The method of claim 1 further comprising generating, in the telephony device, the first  
2     packet signal conforming to a telephony packet protocol and having an audio portion.

1    10.    A audio web telephone system comprising:

2            a telephony gateway in communication with a public switched telephone network  
3     (“PSTN”), the telephone gateway configured to a) receive a circuit-switched signal from a

4        telephony device over the PSTN and b) convert the circuit-switched signal to a telephony packet  
5        protocol signal having an audio portion;  
6                an Internet protocol (“IP”) network;  
7                an audio browser in communication with the telephony gateway to receive the telephony  
8        packet protocol signal and in communication with the IP network..

1        11.      The system of claim 10 wherein the audio browser further comprises:

2                a voice XML browser;  
3                a navigation module;  
4                a content retrieval module; and  
5                a telephony interface module.

1        12.      The system of claim 10 further comprising web cache.

1        13.      The system of claim 11 wherein the navigation module further comprises one of speech  
2        recognition module and touch tone (DTMF) recognition module.

1        14.      The system of claim 11 wherein the content retrieval module further comprises one of  
2        text-to-speech module and streaming media module.